WOCCI 2017: 6th International Workshop on Child Computer Interaction
13 November 2017, Glasgow, Scotland, UK

Visual-feedback in an interactive environment for speech-language therapy

André Grossinho, João Magalhães, Sofia Cavaco

NOVA LINCS, Departamento de Informática
Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa
{jm.magalhaes, scavaco}@fct.unl.pt

Abstract

By combining visual-feedback and motivational elements, a speech therapy computer-based system can offer new approaches with various advantages when compared to traditional speech therapy techniques. Through visual-feedback and adaptation of traditional speech sound exercises, it is possible to create an engaging environment with motivation focused elements. These elements can be used in an interactive environment that motivates the therapy attendee towards better performances. Hereby we present an interactive gamified environment for speech therapy that combines visual-feedback and motivational components. The results from a survey and a usability study suggest that children can show more interest in the speech therapy sessions when the proposed environment is used.

1. Introduction

Speech sound disorders (SSD) of many different types and severities are very common during childhood. As reported by Guimarães et. al [1] for data on European Portuguese (EP), 8.8% of preschool-aged children show SSD. In some cases, the natural development of social skills and the related quality of life of children with SSD may be affected as they struggle to express themselves correctly [2].

These children can (and should) attend speech therapy to address the SSD. In traditional speech therapy a diagnose is made to perceive the real cause of the disorder and from there a speech therapy path is accordingly designed. In subsequent sessions many different exercises that address the identified problem are performed. Many of these exercises rely on auxiliary material, like paper material with drawings and pictures that are used as visual cues for the sounds that must be trained.

Visual-feedback mechanisms such as mirrors are commonly used in speech therapy in exercises for strengthening the face and mouth muscles and in exercises for SSD (figure 1[1]). These allow the child to observe himself and better compare his orofacial expression and motion with the therapist's expression. These visualization perspectives can be explored in exercises by both the therapist and the patient. With the mirror, the speech and language pathologist (SLP) is able to provide active correction examples of orofacial movements and speech articulators posture, while the child is able to observe her own expression and movements, and can more easily try to achieve ideal placement and movement of the vocal tract through successive comparisons. Also self visualization allows the child to understand better the relation between her intentions and actions.

One key aspect for the success of speech therapy is the child's interest and motivation. Faster improvements can be

[1]Image from www.berktree.com.

Figure 1: *Traditional physical layout for speech therapy sessions. A mirror is commonly used in speech-language therapy for bio-feedback.*

reached if the child is motivated and collaborates with the SLP. As reported by several SLPs we interviewed, to keep the child interested in the therapy may be a challenge. Having motivational elements that the child enjoys and looks forward to receiving can contribute to more successful therapy sessions.

Here we propose VisualSpeech, an interactive gamified environment that combines visual-feedback with other visual cues and motivational components for speech therapy for children with SSD. Figure 2 illustrates the new physical layout after replacing the mirror by a visual-feedback hardware/software. By combining visual-feedback with adapted traditional speech sound exercises, it is possible to create an environment with motivation focused elements that can improve children's performance and engagement in speech therapy sessions. The visual cues are used both as support material to illustrate the sound that the child should practice and to give motivation and feedback about the child's performance. The main novelty of this work is the integration of visual-feedback with gamification components. While other computer tools for speech therapy exist, they do not integrate these components [3, 4, 5, 6, 7, 8].

The environment was shown to and validated by SLPs and by a usability study with children. Surveys showed that, according to the SLPs perception, the proposed environment can improve children's performance and engagement during speech therapy sessions. That observation was confirmed in the usability study, in which most children showed more interest in the therapy sessions when the environment was used.

10.21437/WOCCI.2017-6

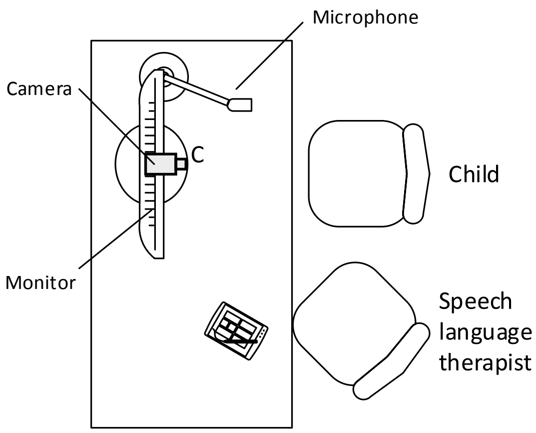

Figure 2: *The physical layout during the speech language therapy session.*

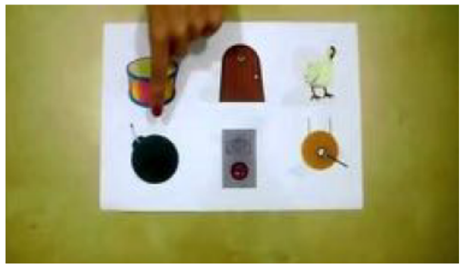

Figure 3: *Visual cues used in tradtitional speech therapy sessions. (Copyright ©2015 Relicário de Sons).*

2. Related Work

A few computer tools that provide different approaches in aiding speech therapy have been proposed and some are already well established [9, 10, 6]. Some more complex systems focus on providing high-quality speech therapy aids and therapy exercises in different forms, with interesting and useful types of visualizations like the Ortho-Logo-Paedia (OLP) and TERAPERS [11, 5]. The Ortho-Logo-Paedia (OLP) provides a vocal tract visualization while the TERAPERS project includes exercises with a 3D model of the vocal tract articulators to be used during the therapy sessions or at home with a PC or PDA.

Other simpler systems, with a less comprehensive approach, target specific aspects of the therapy by providing exercises in a more fun and entertaining way. These include the ARticulation TUtoR (ARTUR) and the Comunica project [6, 8]. The ARTUR system includes exercises that are guided by a virtual tutor, and when needed it provides vocal tract animations. The Comunica framework provides colorful and animated feedback in accordance to the target age. It also provides exercises with images of growing difficulty for improving language understanding of impaired children.

Only a few computer games for speech therapy in EP have been proposed. Among a few others, these include the Interactive Game for Vowel Training [4], in which children can control a car by uttering the vowels in EP, and VITHEA, which is aimed at treating people with aphasia [3]. The VITHEA system uses a virtual tutor to guide patients in naming objects and actions or completing sayings and associations.

All the described systems have the same goal for the in sessions context. They aim at providing a small set of specific exercises that replace a small set of exercises already done in traditional speech therapy without computer aids. While doing this some of them (TERAPERS and ARTUR) provide new useful information and feedback with the use of the 3D vocal tract model. Although useful the systems in the session context lack not only flexibility to both the person in therapy and therapist, but also lack more engaging motivational components.

The exercises provided by the referenced systems are too specific and leave no margin for patient or SLP adaptation. An ideal solution to this lacuna is providing a flexible environment, instead of a specific set exercises, that can be used by the SLPs to perform the exercises with the children as they see fit to themselves and to the children. Our proposal, while oriented for a class of SSD and containing means for specific exercises, is to provide an environment that the SLP can adapt to each patient. Instead of incorporating specific exercises, we propose to include support material that the SLP can use to create exercises.

On the other hand, all the referenced systems present exercises in a more motivational way than traditional therapy by approximating them to a game, presenting them with a virtual tutor or just by being shown in a computer. While these systems are undoubtedly interesting and probably motivating, the motivational-focused elements could be further explored. Since the motivation and interest of the children in therapy is so important in the success of therapy, we find this a major flaw in these systems. The combination of specific not-flexible exercises with lack of motivational-oriented elements can make these systems become repetitive and tedious in the long run. In our game-like environment we give emphasis to the motivational components providing elements with that intent.

3. The gamified interactive environment

VisualSpeech is an interactive environment for speech therapy for children with SSD. While we are currently extending VisualSpeech to include exercises for voice disorders and games to be used at home (for intensive or home training), the work described here focuses on articulation disorders and it is designed to be used during the speech therapy sessions.

Since the learning process of a yet not acquired sound is gradual, traditional speech therapy starts with visual cues that suggest single sounds and later illustrations for full words or sentences. Those single sounds consist of phonemes, the basic unit of speech, and which are produced through different resonances of the air flow originated in the vocal folds of the different shapes of the vocal tract made by its articulators (tongue, lips, nasal cavity, etc.). The exercises of the first stage of speech therapy use visual cues that propose sound productions that consist of or contain the target phoneme such as an already familiar onomatopoeia sound (figure 3). In the subsequent stages of the therapy these cues may suggest whole words or sentences that contain the target sound in all syllabic positions.

VisualSpeech addresses the first stage of speech therapy: phoneme productions. The environment combines visual-feedback with visual cues and motivational components to motivate children to collaborate with the SLP and achieve better speech productions faster than with traditional methods. Below we explore VisualSpeech's components in detail.

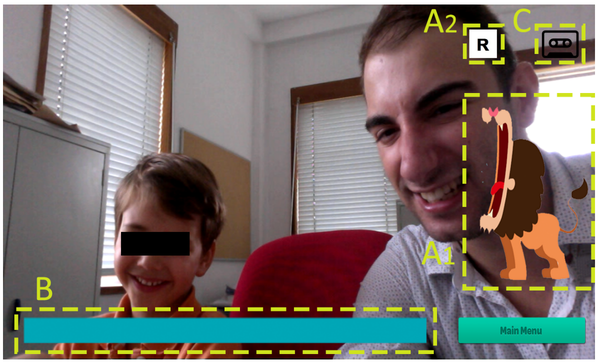

Figure 4: *The proposed environment and its elements. Region A1 contains the visual cue used to illustrate the sound that should be practiced. Region A2 has the phoneme symbol (from the international phonetic alphabet, IPA) for the SLP reference. Region B contains the progress bar. Region C indicates whether video capture is being performed.*

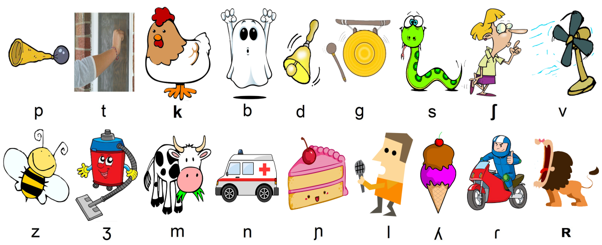

Figure 5: *Visual cues used by VisualSpeech for EP phonemes.*

3.1. Visual-feedback and visual cues

The visual-feedback is achieved through live capture with a webcam, thus substituting the traditional mirror with the computer screen. In order to have both the image of the SLP and the child in the screen, this component occupies as much screen space as possible. This way, the child can observe his own image in the screen but also that of the SLP next to him, which facilitates the comparison of the speech articulators posture (figure 4).

SLPs frequently use visual cues in traditional speech therapy for SSD (figure 3). These cues are illustrations that suggest the speech productions that must be trained. Our environment includes visual cues for phonemes that are controlled by the SLP. Instead of being shown in paper mediums, these are displayed in the screen along with the visual-feedback, which contributes to less distractions as only one illustration is shown each time as opposed to several illustration in the same piece of paper (region A1 in figure 4, and figure 5). While the images used for the visual cues were validated by an SLP, other SLPs can easily change the images used if they do not find them appropriate. The environment includes an indication of the phoneme that is being practiced (region A2 in figure 4). Of course, children (and many adults) may not know these symbols, but this feature is designed specifically for the SLP's convenience. Note that, as mentioned above, the proposed environment was designed to be used during speech therapy sessions.

VisualSpeech uses only the visual cues to suggest the phonemes. No sound is emitted by the environment to explain the phoneme to the child. This was intentional. Since the environment is designed for speech therapy sessions, where the SLPs are present and participate actively in the activities, they will produce the phonemes whenever they find it necessary. Also, the environment can be used to access the child's speech productions, and, as explained to us by several SLPs, in this situation it is preferable that the environment does not play a correct production. In this situation, SLPs want to hear the child's production in their pure form (without having the child trying to copy someone's or the computer's productions).

Combining visual-feedback and visual cues in a single computer environment allows to take advantage of the benefits of using a mirror while the exercises with the visual cues are being performed. Furthermore the use of only one medium (the screen) and the limited positioning and movement associated to it are ideal for observation of the children's vocal tract and steady focus on exercises.

3.2. Speech production capture

Often, SLPs need to analyse the child's productions after the therapy sessions. In these cases they need to record the data during the sessions. To address this need, our environment also provides a way to capture speech production data. While audio recording during the sessions is widely done by SLPs, video recording is not so common. This is mainly due to the destabilizing effect the video capture has on children. When children are aware of the recordings, they get distracted with the procedure, which can affect their speech productions.

Our environment allows to capture audio and image in a discrete way, which avoids the distraction effect. At any moment the SLP can press a key in order to start recording discretely, without the child's notice. To provide a visual confirmation of the recording process to the SLP, a small black tape icon turns red (region C in figure 4). The recordings can later be analysed, for monitoring the child's progress or making better diagnosis. While we have also developed a tool for browsing and searching these recordings, we will not discuss it here [12].

3.3. Progress visualization

While the SLP gives feedback to the child, words of praise and encouragement may not be enough. Our proposal is to include motivational elements that aim at keeping the child motivated and focused in the therapy exercises. In particular, the environment has a performance bar which indicates how well the child is performing in the exercise. The performance bar can be as simple as a plain bar (region B in figure 4) or a funnier bar like an ice cream or a car in a race track (ice cream in figure 6). These elements also provide useful feedback to the child, who can see changes in the environment that depend on his performance. With a bar that increases or decreases according to the child's performance, the child can be encouraged to outperform his last speech productions.

The performance bar also gives the child a visual cue of how far he is from ending the exercise instead of the perspective of indefinite repetitions, which can also motivate the child to try his best in order to reach the end of the bar. In addition, the concept of a bar present in the exercises approximates them to a game, thus increasing the child's interest in the therapy.

In the proposed environment's current version, the performance bar is controlled manually by the SLP through the keyboard. To make it more discrete, this can also be done with the computer mouse. While we do not want to

Figure 6: *Progress bar (ice cream) and rewards (virtual button).*

substitute the SLPs, and the final decision should be theirs, we are investigating ways of making the progress bar advance automatically. Speech recognition techniques can provide valuable feedback to the SLPs, suggesting when to advance the progress bar [13].

3.4. Rewards

When the end of a game is reached, it gives the players a sense of satisfaction. This feeling is what makes players came back to play it again. To get the therapy exercises to deliver this kind of enjoyment we added rewards to the proposed environment.

The goal of these rewards is to provide the final motivation leading to re-playability. They consist of small time limited environment-patient augmented reality fun interactions that the child can play with once she reaches the end of the progress bar. The rewards run for a short time only (5 to 10 seconds), which is ideal for a quick fun interaction without taking the mind off the task at hand.

Currently VisualSpeech includes two types of rewards: a photo boot that adds fun items to the child's face photo and an augmented reality button that shows up in the screen and that drops little colorful balls when triggered (figure 6). When the button appears at a random place on the top of the screen, the child is encourage to reach out her arms and try to press the button. When the button is pressed, colorful particles emerge in a fireworks-like fashion. The pressed status of the button is given by the motion detection in that area of the webcam capture. The motion detection is made by comparison of RGB values within a threshold in the area currently occupied by the button. This reward has the duration of 5 seconds. The short-time of this reward has also the intent of making the child be alert and more energetic. This status can then be readily exploited in the next therapy exercise.

The photo booth reward takes a picture of the child and adds some kind of funny clothing or accessory to the child's face. The added accessory that appears is chosen by the SLP. Many different accessories were included to allow the SLP adapt the reward to the child's gender, age and interests.

This effect is achieved using face detection with the OpenCV detector. The detector uses Viola and Jones' detection method where Haar feature-based cascade classifiers are used [14]. This approach groups face features into different stages of the classifier and applies them one-by-one. If a window passes all stages there is a face recognition. This is applied to each 24×24 window in the input image (in our case the frame from the camera). While this detection is very robust, it can still fail to detect a face in some scenarios. Scenarios that may lead to no detection include: bad lighting; too much face rotation; the person wears thick glasses; or there is too much hair blocking some part of the face.

4. Evaluation

In order to assess if the proposed environment can improve children's motivation and engagement in speech therapy sessions, we planned a usability study to be run during speech therapy sessions with children. Despite our efforts to have many SLPs running this study with children, only three agreed to help us, and from these only one actually ended up collaborating in this test.

To overcome this difficulty, and be able to gather more feedback from SLPs, we ran a survey with SLPs. In order to make sure we were able to collect answers from a sufficient number of SLPs, apart from collecting answers from individual SLPs at their offices, we also run the survey to speech and language therapy graduate students in Escola Superior de Saúde do Alcoitão (a higher education institution). In this way, this time, we managed to collect answers from 26 SLPs. Below we discuss the results obtained from the survey (section 4.1) and the study (section 4.2).

4.1. Validation with SLPs

We validated the proposed game environment with a survey to 26 SLPs to whom we demonstrated the environment. We first examined the background of participating SLPs in terms of used instruments and target population (figure 7). We noted that only 61.5% of the SLPs used software tools and 84.6% used the mirror as a visual-feedback tool. Moreover, the intersection of the SLPs that use the mirror and work with children is 76.9%. All participant SLPs considered that VisualSpeech may be a good replacement of the mirror. Many actually demonstrated interest in using VisualSpeech during their therapy sessions.

A finer grain study was also completed with the SLPs to assess the environment in four different dimensions (figure 8): discreteness, distractedness, performance and engagement. The replacement of the mirror is almost seamless as noted by 88.5% of the therapists. While most SLPs answered that the

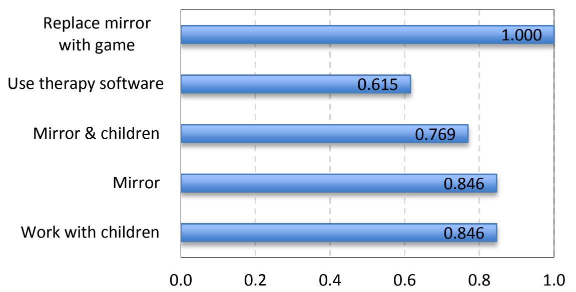

Figure 7: *SLP's practice survey of used instruments and methods, and about VisualSpeech's mirror feature.*

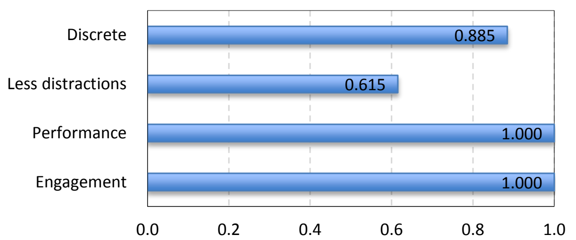

Figure 8: *Survey of speech-interaction games as a replacement of mirror.*

environment will bring less distractions to the sessions, 38.5% did not agree with this statement. This is actually a key point, because it was our goal to make therapy sessions less dull and more interactive. Thus, our goal was to direct children's attention to the speech-exercises. Following this reasoning, it is interesting to observe that all SLPs answered that children's performance is potentially higher with the environment than with a simple mirror. This is probably due to the levels of engagement brought by the game play. Finally 100% of the participating SLPs think that the environment brings positive benefits to speech-language therapy both in terms of performance and engagement.

In general, the SLPs feedback was very positive, with constructive comments about the proposed tool. Some noted that the visual cues were not well adjusted and would prefer to select different images, which is now a configuration option of the environment.

4.2. In-session validation

In order to test our environment in real speech therapy sessions with children, and analyze the reaction of children and SLPs, we ran a usability study. As mentioned above, despite our efforts to have more participating SLPs, there was only one SLP participating in this study. The participant SLP was 23 years old and had 2 years of experience. Although we could only get one SLP to use VisualSpeech, this SLP used it in sessions with 14 children with ages from 4 to 10 (the average age of the participants is 6).

During the study the SLP used the environment as many times as she found appropriate in two separate sessions with each child. After the two sessions the SLP answered a survey in which she scored the child's reaction to the interactive environment and its elements. A five point Likert scale was

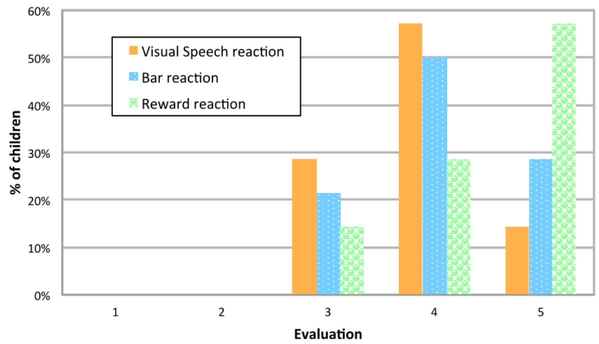

Figure 9: *Children's reaction to VisualSpeech and its motivational elements.*

used in which 1 was *uninterested* and 5 was *enthusiastic*.

Figure 9 shows the results about the children's reaction to the environment and its elements. The first reaction to VisualSpeech was mainly positive (leftmost bar for each score): the SLP used the highest score (5) to rate the reaction of 14% of the children, the second highest score (4) to rate the reaction of 57% of the children and a score of 3 for the remaining children. The reaction to the motivational elements was also positive with children's reactions being scored always above 3. The progress bar had reactions scored always above 3 but mainly with 4 (for 50% of the children). As expected, the rewards caused the most positive reaction with 57% of the children's reactions being scored with 5 and 29% with 4. Only one child lost interest in the performance bar and rewards. The SLP observed that this was probably due to the child's personality and age (10 years old). None of the younger children lost interest in VisualSpeech elements.

It was observed that the children had a better perception of their performance when the progress bar was used, which validates our proposal of using the progress bar as feedback about the speech productions. Also 79% of the children collaborated more with the SLP and obtained better performances when VisualSpeech was used during the sessions.

The SLP gave us very positive feedback about the environment. She felt that the interactive environment was more suited to therapy than the traditional mirror and found the progress bar and rewards very appropriate. Also she did not feel the need to use the traditional visual illustrations when she was using VisualSpeech.

5. Conclusion

While computer tools have ways to make speech therapy exercises more motivating, many fall short in that aspect. In this paper we proposed to bring a gamified environment that gives emphasis to the motivational components into the field of speech and language therapy. The goal of the environment is to direct the attention and motivation of the children to the speech exercises through the use of visual-feedback and other visual stimuli. The main novelty is the integration of visual-feedback with the gamification components. While other computer tools for speech therapy exist, they do not integrate these components. The visual-feedback was achieved by the simulation of a mirror on the computer screen and the inclusion of visual challenges related to the production of speech sounds.

We validated the proposed game environment with a survey to 26 SLPs and a usability study made during speech therapy sessions with 14 children. All participant SLPs agreed that the environment can increase children's motivation and engagement during the speech therapy sessions, which positively affects their performance. Moreover, all SLPs felt that the children's performance is potentially higher with VisualSpeech than with a simple mirror. This is probably due to the levels of engagement brought by the game play.

The SLPs' observation about the increased performance was confirmed by the usability study. During the study, it was observed that the majority of children collaborated more and had better performances when using VisualSpeech. The usability study also showed that the progress bar gives valuable feedback to the children, who could more easily evaluate their performances when a progress bar was used. The survey to SLPs and the usability study with children confirmed that the environment is well suited to its target population, and that it may bring advantages to the therapy sessions that may lead to faster improvements of the patient.

6. Acknowledgments

This work was part of the VisualSpeech and the BioVisualSpeech projects funded by the Portuguese Foundation for Science and Technology and the CMU-Portugal Program (grants CMUP-EPB/TIC/0075/2013 and CMUP-ERI/TIC/0033/2014) and NOVA-LINCS (PEest/UID/CEC/04516/2013).

We thank Isabel Guimarães and Escola Superior de Saúde do Alcoitão for help on setting up the survey, and the SLPs Maria Luís Santos and Diana Lança for their availability and feedback. Thanks also to all SLPs and children who participated in the validation (survey or user test) of this work.

7. References

[1] I. Guimarães, C. Birrento, C. Figueiredo, and C. Flores, "Teste de articulação verbal." Lisboa, Portugal: Oficina Didáctica, 2014.

[2] C. M. Schuele, "The impact of developmental speech and language impairments on the acquisition of literacy skills," *Mental retardation and developmental disabilities research reviews*, vol. 10, no. 3, pp. 176–183, 2004.

[3] A. Abad, A. Pompili, A. Costa, I. Trancoso, J. Fonseca, G. Leal, L. Farrajota, and I. P. Martins, "Automatic word naming recognition for an on-line aphasia treatment system," *Computer Speech & Language*, 2012.

[4] M. Carvalho, "Interactive game for the training of Portuguese vowels," Master's thesis, Faculdade de Engenharia da Universidade do Porto, 2008.

[5] M. Danubianu, S.-G. Pentiuc, O. A. Schipor, M. Nestor, I. Ungureanu, and D. M. Schipor, "TERAPERS-intelligent solution for personalized therapy of speech disorders," *International Journal On Advances in Life Sciences*, vol. 1, no. 1, pp. 26–35, 2009.

[6] O. Engwall, O. Bälter, A. Öster, and H. Kjellström, "Designing the user interface of the computer-based speech training system ARTUR based on early user tests," *Behaviour & Information Technology*, vol. 25, no. 4, pp. 353–365, 2006.

[7] Y. Rybarczyk, J. Fonseca, and R. Martins, "Lisling 3D: a serious game for the treatment of Portuguese aphasic patients," in *Proc. 12th conference of the Association for the Advancement of Assistive Technology in Europe*, 2013.

[8] O. Saz, S.-C. Yin, E. Lleida, R. Rose, C. Vaquero, and W. R. Rodríguez, "Tools and technologies for computer-aided speech and language therapy," *Speech Communication*, vol. 51, no. 10, pp. 948–967, 2009.

[9] F. R. Adams, H. Crepy, D. Jameson, and J. Thatcher, "IBM products for persons with disabilities," in *Global Telecommunications Conference, 1989, and Exhibition. Communications Technology for the 1990s and Beyond. GLOBECOM'89., IEEE*. IEEE, 1989, pp. 980–984.

[10] K. Vicsi, P. Roach, A. Öster, Z. Kacic, P. Barczikay, A. Tantos, F. Csatári, Z. Bakcsi, and A. Sfakianaki, "A multimedia, multilingual teaching and training system for children with speech disorders," *International Journal of speech technology*, vol. 3, no. 3-4, pp. 289–300, 2000.

[11] A. Öster, D. House, A. Protopapas, and A. Hatzis, "Presentation of a new EU project for speech therapy: OLP (Ortho-Logo-Paedia)."

[12] A. Grossinho, S. Cavaco, and J. Magalhães, "An interactive toolset for speech therapy," in *Proceedings of Advances in Computer Entertainment Technology Conference (ACE)*, 2014.

[13] A. Grossinho, I. Guimarães, J. Magalhães, and S. Cavaco, "Robust phoneme recognition for a speech therapy environment," in *Proceedings of IEEE International Conference on Serious Games and Applications for Health (SeGAH)*, 2016.

[14] P. Viola and M. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.